Cascade energy optimization for waste heat recovery in distributed energy systems

Xuan Wang[a,b], Ming Jin[b,c], Wei Feng[b,*], Gequn Shu[a], Hua Tian[a], Youcai Liang[d]

[a] State Key Laboratory of Engines, Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China
[b] Energy Technologies Area, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[c] Industrial Engineering and Operations Research Department, University of California, Berkeley, Berkeley, CA 94720, USA
[d] Systems, Power & Energy Research Division, School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK

A B S T R A C T

The efficiency of distributed energy systems can be significantly increased through waste heat recovery from industry or power generation. The technologies used for this process are typically dependent on the quality and temperature grades of waste heat. To maximize the efficiency of cascade heat utilization, it is important to optimize the choice of waste heat recovery technologies and their operation. In this paper, a detailed mixed integer linear programming optimization model is proposed for waste heat recovery in a district-scale microgrid. The model can distinguish waste heat quality for planning and operation optimization of distributed energy systems. Heat utilization technologies are formulated in this developed model and categorized in different temperature grades. The developed model is validated using four typical cases under different settings of system operation and business models. It is found that the optimization model, by distinguishing waste heat temperature, can increase energy cost savings by around 5%, compared to models that do not consider waste heat temperature grades. Additionally, the results indicate that the developed model can provide more realistic configuration and technologies dispatch.

1. Introduction

A key component of distributed energy systems (DES) is the placement of small-scale energy generation units close to end use loads [1]. DES can help avoid electricity transmission losses, enable flexible dispatch of generation technologies and increase system efficiency, as the electricity generated is used locally and the system incorporates a variety of advanced energy technologies [2]. Issues such as energy efficiency needs, rising energy costs, transmission and distribution infrastructure constraints, and sustainability concerns have increased the attractiveness of DES. Due to the complexity of these systems, which incorporate technologies such as solar photovoltaics (PV), thermal energy storage, power grids, microgrids, and absorption chillers, the optimization of planning and operation are crucial [3]. The choice and operation of waste heat recovery technologies (WHRT) is a key factor in improving energy efficiency of the whole system. Internal combustion engines (ICE) and gas turbines are the most important prime movers in the DES, while their efficiencies are low, ranging from 30% to 50% and 30% to 40%, respectively [4] and resulting in large discharges of energy. In addition to these prime movers, there are many other sources of waste heat that can be reused in a DES. A literature search [5] shows a detailed estimation of the world waste heat sources in 2012, accounting for about 71.8% of primary energy input. A number of research efforts such as [6] have proven that with appropriate WHRTs, the primary

* Corresponding author.
E-mail addresses: weifeng@lbl.gov (W. Feng), sgq@tju.edu.cn (G. Shu).

Nomenclature

T	temperature (°C)
m	mass flow rate (kg/h)
C_p	specific heat (kJ/kg K)
Q	absorbed heat
QC	cooling (kWh)
H	heating (kWh)
P	power (kWh)
C	cost
c	unit cost
CP	capacity (kW)
NG	nature gas heat (kWh)
Z	temperature binary
X	install binary
b	mass flow rate binary
η	efficiency

Subscript

inv	investment
op	operation
mtn	maintenance
pur	purchase
sal	sale
var	variable
rat	rated
m	month
d	day
h	hour
ex	exhaust

Acronyms

WHRT	Waste Heat Recovery Technology
ICE	Internal Combustion Engine
CCHP	Combined Cooling Heating and Power
CEU	Cascade Energy Utilization
DES	Distributed Energy System
MINLP	Mixed Integer Non-Linear Program
MILP	Mixed Integer Linear Program
RC	Rankine Cycle
ORC	Organic Rankine Cycle
DARS	Double-effect Absorption Refrigeration System
AHP	Absorption Heat Pump
ARS	Absorption Refrigeration System
DH	Direct Heating
JW	Jacket Water
NGC	Natural Gas Chiller
EC	Electric Chiller
NGB	Nature Gas Boiler
EB	Electric Boiler
HP	Heat Pump
ST	Solar Thermal
PV	Photovoltaic Generation
EST	Electrical Storage Battery
HST	Heat Storage Tank
CST	Cooling Storage Tank
DR	Demand Response
COP	Coefficient of performance
HT	Hight temperature
LT	Low temperature
VCHP	Vapor compression heat pump

energy ratio of cogeneration system can exceed 80%. As a result, researching WHRTs with respect to the optimization issue in DES is an important potential source of energy savings.

In this work, cascade energy utilization requires thermal-driven technologies to be operated within its own suitable temperature range of the heat source. In other words, the configuration and operation of WHRTs in DES should be based on energy equality. Proper configuration and operation of these technologies, to match the energy quality needed, can greatly improve DES efficiency. Without optimal system design and operation strategies, these technologies can result in even higher energy losses than a traditional energy system [4]. Additionally, heat source temperatures that are too low to drive particular technologies, such as the Rankine Cycle, may lead to unreasonable optimization results for DES in real applications. Consequently, it is crucial to optimize planning and operation of different WHRTs according to energy quality.

2. Literature review

Many existing literatures have studied DES optimization problems. The traditional DES optimization problem can be described simply as: under constraints of load demand and equipment models, use a certain algorithm to find the optimal solution for planning and operation of all the technologies. In short, there are four key aspects to look at in this process: the technology model, the objective function, the calculation method, and load demand. Former studies usually focused on these four aspects, as described below.

Technology modeling is a critical part of optimization. It can be thermodynamic-based, correlation-based, or use an assumed constant efficiency. Thermodynamic-based models are usually complex, so they are limited to simple or well-studied system configurations [7]. Barelli [8] established a thermodynamic model to observe the full functioning of a DES under variable loads and defined the best operation condition in terms of humidity percentage. Lecompte [9] proposed a detailed off-design model of a DES, which allowed for selecting the optimal size of the heat exchangers, the optimal mass flow rates, and other variables. By contrast, complex systems can be optimized with correlation-based models at the sacrifice of accuracies. Nonlinear performance curves describing the off-design behavior of different technologies are the most common method in many research efforts. For instance, one study shows [10] that the efficiencies of internal combustion engine, boiler, turbine, and pump were computed as a function of load in the DES optimization model. Alvarado et al. [11] applied real manufacturer data sheets for the technologies to represent their off-design performance in the optimization model, which was named “data-driven structure.” There are many research efforts in which most component models are assumed to be a product of input energy and constant efficiency. For example, Stoppato et al. [12] invested a model for optimal design and management of a cogeneration system in which most of the equipment efficiencies were constant while they focused on the storage modules and considered battery lifetime optimization.

The objective function is also significant in planning and operations optimization for DES—this function determines whether to add a technology, what the equipment size is, and how to operate various

kinds of equipment. Past research efforts show that there are four main types of objective functions [13]: energy [14], exergy [15], economy [16], and environment [17]. The objective function can also be formulated for single-objective and multiple-objectives with weights that can be defined by the method of "a priori articulation of preferences" or "a posteriori articulation of preferences" [18].

Because of the nonlinear model of the component and complex relationship among optimal optimization variables, a DES optimization problem is often modeled using mixed integer non-linear programming (MINLP). Diverse approaches have been applied to solve a problem such as genetic algorithm [19], simulated annealing [20], particle swarm optimization [21], dynamic programming [22], artificial neural networks [23], and evolutionary algorithms [24]. The large number of integer variables and the nonlinear constraints, which originate when considering multiple interconnected technologies and a long-time horizon (such as one year with an hour-basis discretization), result in an extremely challenging MINLP. Ommen et al. [25] conducted an empirical comparison of the linear programming (LP), mixed integer linear programming (MILP), and non-linear programming (NLP) methods. The research concluded that MILP was the most appropriate model from the viewpoint of accuracy and run time, using the linearization approach to approximately model non-linear technologies. Therefore, many approaches have been proposed to convert the original MINLP into an approximated MILP. The paper [26] showed that the MILP model can be effective for DES optimization with more than 10,000 variables.

Active demand response (DR), defined as "changes in energy usage implemented directly or indirectly by end-use customers/prosumers from their current consumption injection patterns in response to certain signals" [27], has been implemented to reduce peak energy usage and utility bills [28] as well as to bring higher capacity factors and grids security [29]. Owing to the challenges of DR, together with the uncertainty inherent in renewables and wholesale market prices [30], increasing efforts have researched applying DR to the DES optimization process. In work done formerly by the authors of this paper [30], an extremely detailed review of DR was described and a specific DES optimal model with DR was established. The case studied at a campus indicated that 17.5% peak load reduction and 8.8% cost reduction could be achieved with the DR model compared to the non-trivial baseline.

Overall, the research reported in the literature does not consider variations in the energy quality of heat source. Waste heat is simply calculated as the same energy quality and recovered to supply heating or cooling by an assigned technology. Considering waste heat utiliza-tion, there are three primary problem when treating waste heat as a homogenous energy thermal quality in the process of DES optimization. Firstly, if energy quality is much higher than the requirement of the technology, it leads to waste of high quality energy and decreases the efficiency of the whole DES (described as "energy degrading utiliza-tion" in this work). Secondly, if the temperature of the heat source is lower than the minimum temperature limit of certain technologies, it is not feasible to use the heat to drive the equipment—this is described as unreasonable "energy upgrade utilization." Thirdly, owing to the former two problems, a reasonable and detailed dispatch of the waste heat resource could not be optimized. Consequently, it is critical to consider energy quality when optimizing the planning and operation of DES.

Different temperatures of heat are suitable for different WHRTs. This is especially important for waste heat recovery from multiple heat sources of different energy qualities and for high-temperature heat sources with a significant temperature drop after heat recovery. For example, the ICE produces low-temperature jack water (about 90 ˚C) [31] and high-temperature exhaust (about 400–800 ˚C) [32]. As a result, new approaches focused on cascade utilization based on thermodynamic analysis are proposed. Lin et al. [33] made a detailed description of the principle "temperature match and cascade energy utilization" and show several typical examples. Yang et al. [34] proposed a high-efficiency double-effect absorption cycle with engine jacket water as the heat source for a low-pressure generator and exhaust as the heat source for a high-pressure generator in a tri-generation application. Mohammadi [35] proposed a novel cogeneration system, in which the waste heat of the engine exhaust was recovered by the organic rankine cycle (ORC) and the absorption chiller in sequence, because the ORC can use medium-temperature heat to produce elec-tricity of high energy quality, and the absorption chiller is more suitable for low-temperature heat. Hajabdollahi [19] investigated the effects of load demands on the selection of optimum waste heat recovery tech-nologies. The waste heat of ICE exhaust could be recovered by ORC, the absorption chiller, and direct heating in order of energy quality levels; selection of the three technologies was based on load demand. Han et al. [36] proposed a novel high-efficiency hybrid absorption-com-pression refrigeration system based on cascade waste heat utilization. The waste heat was used to heat turbine vapor and then the solution in the absorption refrigeration system in a series. Jing and Zheng [37] also used waste heat to drive a turbine and absorption chiller. Chen et al.[38] used the high-temperature portion of engine exhaust to drive thermoelectric generators and the low-temperature portion to drive a waste heat boiler.

However, the existing literature which focuses on waste heat cascade utilization as mentioned above only performed the thermodynamic analysis of a certain specified DES, and did not optimize planning and operation of the DES. When considering DES optimization, both energy equality and load demand should be considered at the same time, making the problem more complicated. Related DES optimization research with cascade utilization is quite rare. Bischi et al.[39] divided the thermal energy to high-temperature flow and low-temperature flow in a MILP model for tackling the short-term scheduling problem. Gao et al. [40] proposed a methodology of energy matching and optimization for a DES based on energy level analysis. However, in Bischi et al.'s research [39] only two kinds of WHRTs were used and only two energy levels were divided, which could not distinguish all of the suitable temperature ranges for different WHRTs. Furthermore, this work only optimized DES operation, and did not optimize planning. The methodology proposed by Gao et al. [40] is only suitable for operation optimization of an already-designed combined cooling, heating, and power system (CCHP).

In conclusion, there is a need to develop a detailed optimization model for planning and operation of a DES by distinguishing input waste heat quality. Hence, this study proposes a novel waste heat uti-lization optimization model. In the model, the energy quality level of the heat source is distinguished by temperature. Detailed temperature thresholds for almost all common WHRTs are worked into the con-straints of their models. The optimization model can select the most suitable WHRTs based on energy quality and load demand while pro-viding detailed operation dispatch for related energy equipment. Four cases are investigated to show the effectiveness of this model. The calculated results are compared with the model without distinguishing energy quality in the base case. Based on the literature review for this section, the model developed in this study can make the following original contributions regarding DES:

1. A novel cascade energy optimization model, able to distinguish energy quality, is developed for use in DES planning and operation.

The model allows users to choose appropriate WHRTs based on energy quality and load demand.
2. The model avoids unnecessary heat upgrade utilization and efficiency loss by utilizing waste energy in optimal temperature ranges.
3. Cascade energy utilization provides detailed optimal operation for heat utilization in different temperature ranges. Results show a significant increase in system efficiency.

In summary, the model developed in this study provides an optimal method for waste-heat utilization strategy, delivering reasonable, efficient and economical system planning and operation for a DES.

3. Energy equality matching for waste heat recovery technologies

As mentioned above, the work featured in this paper focuses on optimizing the planning and operation of WHRTs based on energy quality and load demand. The WHRT is a type of "passive" technology. Unlike "active" technologies such as natural gas boilers or electrical chillers, passive technologies rely on heat source energy quality to determine whether the technology can be operated/used and energy quality has a significant effect on energy efficiency. Temperature is an important parameter to distinguish energy quality. Section 3.1 describes suitable temperature ranges for common technologies.

3.1. Temperature ranges of common technologies

3.1.1. Rankine cycle and organic rankine cycle

Rankine cycle (RC) is a mature technology and has long been used to recover high temperature exhaust heat for producing electricity [41]. A combined system using a gas turbine and RC can reach an efficiency of approximately 60% [42]. This system uses water as a working fluid, and the water needs to achieve a large superheat degree before entering the turbine, so the heat source temperature is required to be relatively high, usually higher than 400 °C [43].

The organic rankine cycle (ORC) operates similarly to the conventional steam RC except it uses organic working fluid instead of water. Efficiency in the ORC is approximately 5–20% [44]. ORC is the best technology to convert heat into power when the temperature source is in the range of 200–400 °C [45]. A variety of organic fluids can be used in an ORC and they have different efficiencies. In literature [46], working fluids are dived into high temperature (HT) and low temperature (LT) working fluid. LT working fluids can use low temperature heat (100–200 °C), resulting in efficiencies of about 5–10%. HT working fluids are more suitable for higher temperatures and have higher efficiencies (10–20%). Therefore, this paper divides ORC into HT and LT.

3.1.2. Absorption refrigeration system

The absorption refrigeration system (ARS) is one of the most commonly applied and commercialized cooling WHRTs in DES. There are single-, double-, and triple-effect absorption refrigeration systems; temperature requirement increase respectively. Triple-effect systems have not been applied widely in the market [47]. Table 1 shows the temperature range for single- and double-effect ARS.

3.1.3. Vapor compression heat pump

The vapor compression heat pump (VCHP) can achieve temperature lifts with significant consumption of electrical energy. The heat-source temperature of the VCHP is usually lower than 30 °C, because high temperatures increase gas temperatures before and after the compressor, affecting the safety of the compressor [48].

3.1.4. Absorption heat pump

Absorption heat pump technology (AHP) replaces electrical consumption with heat energy to achieve its effect. It can be divided into the type-one absorption heat pump (1AHP) and type-two absorption heat pump (2AHP). The former is also called a "heat increasing HP" and uses high temperature heat to produce a large amount of middle temperature heat. 2AHP is also called a "temperature-rising HP" or "absorption heat transformer." This technology can convert low-temperature heat to middle-temperature heat with a lift of around 30–50 °C by a single stage HP [50]. Table 1 shows suitable temperature ranges for various HPs.

3.1.5. Direct heating (DH)

Low-temperature waste heat can be directly integrated into conventional district heating networks. For example, average supply temperatures are 86 °C in Sweden [47] and 70–120 °C in Denmark [53]. In China, the literature showed three specifications of 95 °C/70 °C, 85 °C/65 °C, and 75 °C/55° [49]. Next generation low-temperature networks are developing in order to utilize low-temperature heat.

3.2. Technology summary

Table 1 gives a summary of suitable temperature ranges for common technologies and their performance coefficients. It should be noted that the temperature ranges in this table are approximate, because the lower limit temperatures are greatly affected by equipment and the form of heat source. Fig. 1 shows an energy flow tower that describes the energy quality levels of common waste heat sources used in DES with corresponding suitable technologies. In the energy flow tower, technologies are aligned with specific temperature ranges. The high-temperature range can reduce its energy level to drive the low-level technologies with exergy loss if necessary, but not the other way around, according to the second law of thermodynamics. Temperature constraints in Fig. 1 are included in the optimization model presented in Section 4.

There may be several kinds of heat sources in an optimization problem of DES, such as the industry waste heat, condensing heat from a power plant, jacket water, and exhaust from the primary mover (shown in Fig. 1). These heat sources can be divided into two classes. The first class produces a large temperature drop after heat recovery, such as exhaust. This kind of waste heat enables cascade utilization using a variety of technologies according to energy quality levels and load demand. Section 4 outlines these models as shown in the energy flow tower (Fig. 1).

The second class of heat sources has only a small temperature drop after heat recovery, such as jacket water. Jacket water temperatures out of the engine are required to be around 90 °C; when returned to the engine, temperatures must be about 70 °C [13]. As a result, this kind of

Table 1
Suitable temperature ranges for common technologies.

Technology	Temperature	Performance coefficient	Application
VCHP [48]	< 30 °C	COP: 3–5	Heating
DH [49]	> 60 °C	Efficiency > 0.8	Heating
2AHP [50]	> 60–80 °C	COP: 0.4–0.5	Heating
ARS [51]	70–170 °C	COP: 0.4–0.8	Cooling
1AHP [52]	> 120–240 °C	COP: 1.6–1.8	Heating
DARS [52]	> 140–240 °C	COP: 0.8–1.2	Cooling
ORC [45–46]	100–400 °C	Efficiency: 0.05–0.20	Electricity
RC [43]	> 400 °C	Efficiency: 20–30%	Electricity

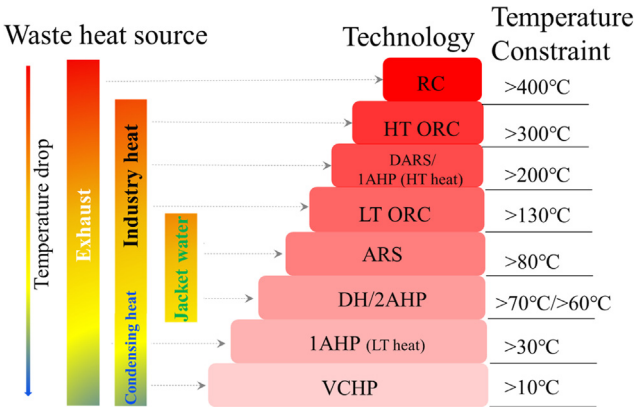

Fig. 1. Energy flow tower of waste heat recovery technologies.

heat source has a narrow interval of energy quality and the candidate WHRTs can be regarded in the same level of the energy flow tower. Industry waste heat of high temperature can be grouped into class 1, while that of low temperature and the condensing heat from power plants can be grouped into class 2.

In this study, the prime mover is an ICE with the waste heat source of exhaust and jacket water. Consequently, the two classes are represented by these two kinds of waste heat source in the study. Other kinds of waste heat source do not appear. Nevertheless, they can be added using the same model of exhaust and jacket water conveniently.

4. Cascade energy optimization model

This model is aimed at optimizing planning and operation of the DES, primarily focused on WHRTs. The optimal plan is the one that minimizes discounted present value of all costs in the planning period (i.e., operating cost plus investment costs). The optimization problem for operational analysis is formulated as a mixed integer linear programming problem and solved with Gurobi. This model can provide alternative decisions to determine which technology should be installed in the DES, capacity of the adopted technologies, and how much energy should be produced by each technology every hour. System operations for one year into the future are considered. The typical time step in the operational model is one hour. The load days represent twelve months with two representative days within the week (weekday and weekend). As a result, total time steps are 576 ($12 * 2 * 24$).

The prime mover is the ICE, so there is exhaust and jacket water waste heat to represent the two classes of heat source as mentioned above in Section 3.2. The discharged exhaust temperature after heat recovery is usually required to be higher than the acid dew point, otherwise the equipment can be corroded [54]. The acid dew point is determined by the sulphur content of the fuel, and it is assumed to be 100 °C in the study, which means the exhaust heat below 100 °C is discharged. Therefore, the WHRT candidates are RC, ORC, double-effect ARS (DARS), single effect ARS, 1AHP, and DH. Since the jacket water temperature is low and its recoverable temperature range is small, the candidates are limited to ARS and DH, which are named JWARS and JWDH in order to differ from exhaust. The active candidate technologies include: natural gas chiller (NGC), electric chiller (EC), natural gas boiler (NGB), electric boiler (EB), heat pump (HP), solar thermal (ST), PV, and ICE. Storage technologies include electrical storage battery (EST), heat storage tank (HST), and cooling storage tank (CST). Their technology parameters such as efficiency and cost are the initial parameters set by the users. The load demand of electricity, heating, and cooling can be satisfied by these technologies as shown in Fig. 2. The mathematical models are described step by step in detail by presenting the objective function and all the constraints.

4.1. Objective function

The goal of optimization is to minimize overall annualized investment cost and annual operating cost of the whole energy system while

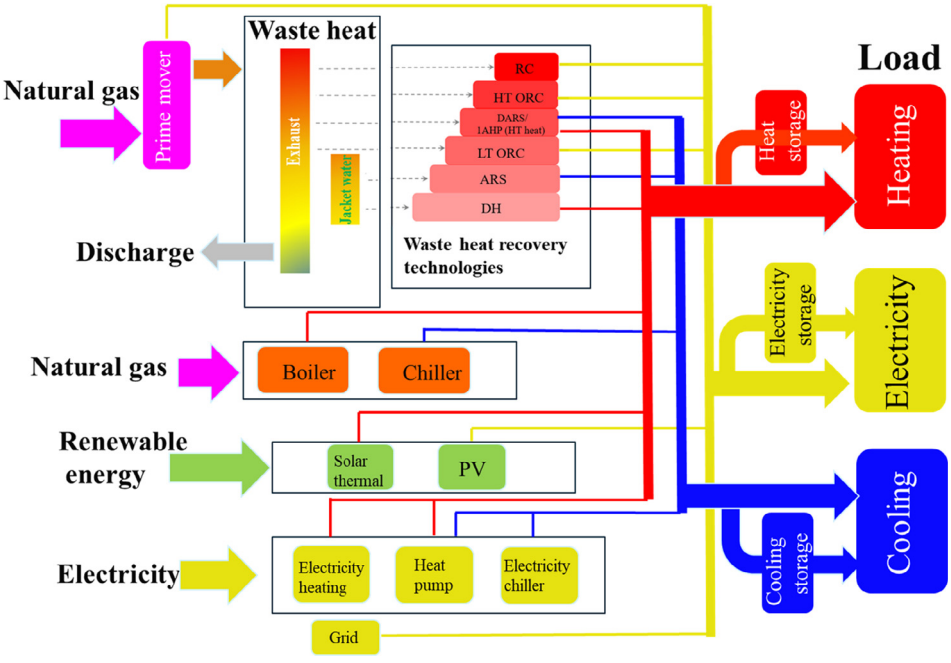

Fig. 2. Schematic representation of the DES.

maximizing profit by selling surplus electricity and heating energy outside the system:

$$\min C_{\text{cost}} = C_{inv} + C_{op} + C_{mtn} + C_{pur}^{grid} + C_{pur}^{heat} - C_{sal}^{grid} - C_{sal}^{heat} \tag{1}$$

where C_{inv}, C_{op}, C_{mtn} are investment cost, operation cost, and maintenance cost, respectively. C_{pur}^{grid} and C_{pur}^{heat} are the purchase cost for electricity and heat. C_{sal}^{grid} and C_{sal}^{heat} are sale profit of electricity and heat. The detailed calculation of every item in Eq. (1) is conventional and can be found in a literature search [30], so it is not described here.

4.2. Constraints

4.2.1. Energy balance

The DES must satisfy customers' requirements for electricity, heating, and cooling. Electricity loads can be met by PV arrays, the prime mover, Rankine Cycle technology for waste heat recovery, batteries, or grid purchases. Excess electricity produced by the district energy system is considered as income.

$$\begin{aligned} &D_{m,d,h}^{ele\_only} + P_{m,d,h}^{grid\_sal} + P_{m,d,h}^{EC} + P_{m,d,h}^{EB} + P_{m,d,h}^{HP} + P_{m,d,h}^{EST\_sto} \\ &= P_{m,d,h}^{grid\_pur} + P_{m,d,h}^{PV} + P_{m,d,h}^{RC} + P_{m,d,h}^{HTORC} + P_{m,d,h}^{LTORC} + P_{m,d,h}^{PRI} + P_{m,d,h}^{EST\_from} \end{aligned} \tag{2}$$

Heat loads can be satisfied by the natural gas boiler, electrical boiler, solar thermal, heat pump, WHRTs of 1AHP and DH, and heat storage.

$$\begin{aligned} D_{m,d,h}^{heat} + H_{m,d,h}^{sal} + H_{m,d,h}^{HST\_sto} = {} & H_{m,d,h}^{NGB} + H_{m,d,h}^{EB} + H_{m,d,h}^{HP} + H_{m,d,h}^{ST} + H_{m,d,h}^{1AHP} \\ & + H_{m,d,h}^{DH} + H_{m,d,h}^{JWDH} + + H_{m,d,h}^{HST\_from} \end{aligned} \tag{3}$$

Cooling loads can be supplied by the electric chiller, natural gas chiller, WHRTs including DARS, ARS, and cooling storage:

$$\begin{aligned} D_{m,d,h}^{cooling} + QC_{m,d,h}^{CST\_sto} = {} & QC_{m,d,h}^{NGC} + QC_{m,d,h}^{EC} + QC_{m,d,h}^{HP} + QC_{m,d,h}^{DARS} + QC_{m,d,h}^{ARS} \\ & + QC_{m,d,h}^{JWARS} + QC_{m,d,h}^{CST\_from} \end{aligned} \tag{4}$$

4.2.2. Technology model

This study is primarily focused on the optimization of WHRTs, so their models are stated in detail here. Models for all active technologies and storage technologies are reported in past work [30] and so are not stated in detail here.

4.2.2.1. Constraints for exhaust waste heat recovery. The input energy of the exhaust WHRT is calculated as:

$$Q^{\text{ex}} = (T_{in}^{ex} - T_{out}^{ex}) m^{ex} C_p^{ex} \tag{5}$$

Although the exhaust temperature out of engine (T_{in}^{ex}) varies with engine working conditions, it does not change much above the 40% condition [55]. Consequently, in order to reduce the computation task, the exhaust temperature is assumed as a constant value. The exhaust mass flow rate is strongly associated with the input natural gas. According to the experimental research on ICE [55], the relationship between exhaust mass flow rate and input natural gas is approximately linear. Therefore, it is defined as:

$$m_{m,d,h}^{PRI} = \kappa_{gas-exh} NG_{m,d,h}^{PRI} \tag{6}$$

where $k_{gas\text{-}exh}$ is a constant coefficient and assumed as 0.000563 according to the experimental research [55]. The unit used for consumed heat $NG_{m,d,h}^{PRI}$ is kWh.

Exhaust is the main waste heat source used in the model and has a large recoverable temperature range, so it can be recovered with a variety of technologies based on energy quality levels, as shown in the energy flow tower of Fig. 1. Exhaust can flow through the energy flow tower and according to the energy quality and load demand, which technologies should be used and how much energy should be recovered by each technology can be optimized. The highest level in the tower is the RC. Its absorbed heat and the generated electricity are calculated as:

$$\mathrm{Q}_{m,d,h}^{RC} = Z_{m,d,h}^{RC} (T_{m,d,h}^{ex1} - T_{m,d,h}^{ex1'}) m_{m,d,h}^{ex} C_p^{ex} \tag{7}$$

$$P_{m,d,h}^{RC} = \mathrm{Q}_{m,d,h}^{RC} \eta^{RC} \tag{8}$$

Therein, $T_{m,d,h}^{ex1}$ is the exhaust temperature at the RC inlet and $T_{m,d,h}^{ex1'}$ is that at the outlet. Both values are optimal variables for operation: η^{RC} is RC efficiency and $m_{m,d,h}^{ex}$ is associated with input natural gas as mentioned above, so it is also a variable, leading to nonlinearity of Eq. (7). The equation is linearized after the description of the model. $Z_{m,d,h}^{RC}$ is a binary variable that determines whether the exhaust temperature is high enough to drive the RC.

$$\begin{cases} Z_{m,d,h}^{RC} \in [0, 1] \\ T^{ex1} \geqslant 400 Z_{m,d,h}^{RC} \\ T^{ex1} \leqslant Z_{m,d,h}^{RC} (T^{ex} - 400) + 400 \end{cases} \tag{9}$$

The lower limit temperature to drive the RC is assumed to be 400 °C. It can be seen from Eq. (9) that if the inlet exhaust temperature is lower than 400 °C, $Z_{m,d,h}^{RC}$ is 0, and the absorbed heat is 0 (as well as the RC output). Otherwise, $Z_{m,d,h}^{RC}$ is 1 which means $T_{m,d,h}^{ex}$ (the exhaust temperature out of the engine) exceeds 400 °C. Most WHRTs can only utilize part of the exhaust heat, so the outlet exhaust temperature cannot be lower than a certain temperature, which depends on the technical principle and equipment performance. For RC, research shows that the lowest outlet temperature can be assumed as 150 °C [56].

$$T_{m,d,h}^{CHP,ex} \geqslant T_{m,d,h}^{CHP,ex1} \geqslant T_{m,d,h}^{CHP,ex1'} \geqslant 150 Z_{m,d,h}^{CHP,RC} \tag{10}$$

Additionally, there are installment constraints for RC:

$$CP_{\min}^{RC} X^{RC} \leqslant CP_{rat}^{RC} \leqslant CP_{\max}^{RC} X^{RC} \tag{11}$$

$$0 \leqslant P_{m,d,h}^{RC} \leqslant Cp_{rat}^{RC} \tag{12}$$

In the next energy level of HT ORC, the output power is calculated as:

$$\mathrm{Q}_{m,d,h}^{HTORC} = Z_{m,d,h}^{HTORC} (T_{m,d,h}^{ex2} - T_{m,d,h}^{ex2'}) m_{m,d,h}^{ex} C_p^{ex} \tag{13}$$

$$P_{m,d,h}^{HTORC} = \mathrm{Q}_{m,d,h}^{HTORC} \eta^{HTORC} \tag{14}$$

The constrains of $Z_{m,d,h}^{HTORC}$ are:

$$\begin{cases} Z_{m,d,h}^{HTORC} \in [0, 1] \\ T_{m,d,h}^{ex2} \geqslant 300 Z_{m,d,h}^{HTORC} \\ T_{m,d,h}^{ex2} \leqslant Z_{m,d,h}^{HTORC} (T_{m,d,h}^{ex1'} - 300) + 300 \end{cases} \tag{15}$$

$$T_{m,d,h}^{ex1'} \geqslant T_{m,d,h}^{ex2} \geqslant T_{m,d,h}^{ex2'} \geqslant 100 Z_{m,d,h}^{HTORC} \tag{16}$$

The inlet temperature of HT ORC must be lower than $T_{m,d,h}^{ex1'}$ as shown in Eq. (16). If there is RC in the high energy level and its outlet temperature $T_{m,d,h}^{ex1'}$ is higher than the lower limit of HT ORC (300 °C), $Z_{m,d,h}^{HTORC}$ can be equal to 1. By contrast, if $T_{m,d,h}^{ex1'}$is lower than 300 °C, $Z_{m,d,h}^{HTORC}$ is 0, which means the remaining exhaust heat cannot drive the HT ORC and the output is 0. On the other hand, if the RC is not set at "installed" or "operate," $T_{m,d,h}^{ex1'}$is optimized to be equal to $T_{m,d,h}^{ex1'}$ and the HT ORC becomes the first stage.Instalment constraints are:

$$CP_{\min}^{HTORC} X^{HTORC} \leqslant CP_{rat}^{HTORC} \leqslant CP_{\max}^{HTORC} X^{HTORC} \tag{17}$$

$$0 \leqslant P_{m,d,h}^{HTORC} \leqslant CP_{rat}^{HTORC} \tag{18}$$

The models for the remaining stages of WHRT are set up according to the energy flow tower concept and similar to the two models described previously. Therefore, they are not stated in detail here and their models are attached in the Appendix. The energy flow tower concept can be used not only with exhaust, but also with other waste heat sources with large temperature gradients. For example, if the problem includes another high temperature industry waste heat, it can be added with the same model of exhaust, based on the energy flow tower, to be optimized together.

4.2.2.2. Constraints for jacket water waste heat recovery. Jacket water is an important part of engine waste heat and usually accounts for about 15–25% of input energy [20]. The proportion is assumed to be 20% in the model.

$$Q^{JW}_{m,d,h} = 0.2NG^{PRI}_{m,d,h} \tag{19}$$

Jacket water temperature is low and the recoverable temperature difference is approximately 20 °C. Jacket water temperature must be approximately 70 °C at the inlet of an ICE and around 90 °C at the outlet [20]. Therefore, it can be only utilized by DH or ARS.

$$H^{JWDH}_{m,d,h} = Q^{JWDH}_{m,d,h}\eta_{JWDH} \tag{20}$$

$$QC^{JWARS}_{m,d,h} = Q^{JWARS}_{m,d,h}COP_{JWARS} \tag{21}$$

$Q^{JWDH}_{m,d,h}$ and $Q^{JWARS}_{m,d,h}$ are the absorbed heat by JWDH and JWARS, respectively. Their sum must be less than or equal to the amount of jacket water waste heat:

$$Q^{JWDH}_{m,d,h} + Q^{JWARS}_{m,d,h} \leqslant Q^{JW}_{m,d,h} \tag{22}$$

The installation constraints of JWDH and JWARS are:

$$Cp^{JWDH}_{\min}X^{JWDH} \leqslant Cp^{JWDH}_{rat} \leqslant Cp^{JWDH}_{\max}X^{JWDH} \tag{23}$$

$$Cp^{JWARS}_{\min}X^{JWARS} \leqslant Cp^{JWARS}_{rat} \leqslant Cp^{JWARS}_{\max}X^{JWARS} \tag{24}$$

4.3. Linearization

In order to linearize the constraints of energy balance for exhaust WHRTs, the exhaust mass flow rate is discretely uniformly:

$$m^{ex}_{m,d,h} = b_1m_e + b_2m_e + b_3m_e + \ldots + b_nm_e \tag{25}$$

$$b_1 = \begin{cases} 1, & m^{ex}_{m,d,h} \geqslant m_e \\ 0, & m^{ex}_{m,d,h} < m_e \end{cases}; b_2 = \begin{cases} 1, & m^{ex}_{m,d,h} \geqslant 2m_e \\ 0, & m^{ex}_{m,d,h} < 2m_e \end{cases}; b_3 = \begin{cases} 1, & m^{ex}_{m,d,h} \geqslant 3m_e \\ 0, & m^{ex}_{m,d,h} < 3m_e \end{cases}; \ldots b_n = \begin{cases} 1, & m^{ex}_{m,d,h} \geqslant nm_e \\ 0, & m^{ex}_{m,d,h} < nm_e \end{cases}; \tag{26}$$

m_e is the unit discrete value. n is the number of discrete sections. The larger n is, the more precise optimization results are. However, a large n value brings more binary variables, increasing the complexity of the calculation task exponentially. After linearization, Eq. (7) becomes:

$$Q^{RC}_{m,d,h} = C^{ex}_p m_e Z^{RC}_{m,d,h}(b_1 + b_2 + b_3 + \cdots + b_n)(T^{ex1}_{m,d,h} - T^{ex1'}_{m,d,h}) \tag{27}$$

5. Case study

In this section, four cases are selected to study the optimization model. All of the study cases have the candidate technologies shown in Fig. 2, except special mention. For the sake of simplicity, the same load profile and operative cost assumptions are applied for all the cases, except special statement. Case A is the base case and compares to the optimization results with the traditional CCHP model such as in [30,57], which cannot distinguish energy quality. In order to research the optimization results under different business models, the heat cannot be sold or purchased in Case B. In Case C, the lower temperature limit of the RC is assumed to be higher than the exhaust temperature, aiming at testing the function of screening out technologies based on energy quality. The other initial conditions are the same as the base case. In Case D, the investment cost for the heat tank is assumed to be cheap in order to include energy storage modules in optimization results and use up the waste heat. The differences among the initial conditions of the four cases are listed in Table 2. All of the calculation gaps in the four cases are 10^{-3} and there are about 88,000 variables in the model.

The demand profile, with one-hour time steps for the four cases, is represented in Fig. 3, which is cited from the openEI website. As mentioned above, the load days represent 12 months and two representative days within the week (weekday and weekend), so total time steps are 576 ($12 * 2 * 24$) for one year. In order to analyze the operation of the different technologies in detail, two typical days are selected in all the below cases. Typical Day 1 represents the heating season and typical Day 2 represents the cooling season. The main operative cost assumptions and performance parameters of technologies are listed in Tables 3 and 4, respectively.

5.1. Case A (base case)

This is the base case and compares to the optimization results from the traditional CCHP model. In traditional CCHP models such as [13–16,19], waste heat from the prime mover is calculated as the same energy quality and recovered to supply heating or cooling with an assigned technology. The prime mover and WHRTs are usually packed together as one module. By contrast, in this model the prime mover and WHRTs are separated, which enables optimization of the selection and operation of different WHRTs according to energy quality and load demand. For the sake of convenience, it is named the cascade energy utilization (CEU) model in this paper. In order to compare the two models, the traditional CCHP model is set up in the CEU model by replacing the part of WHRTs and the initial conditions of the two models are the same, except for the waste heat recovery portions of the calculation. In the CEU model, the waste heat is from jacket water and exhaust. Jacket water waste heat is assumed to be 20% of input energy and available exhaust heat is assumed to be 23.2% of input energy. Therefore, the available waste heat in the CCHP model is assumed to be 43.2% of input energy and can be recovered to produce heating or cooling. In the traditional CCHP model, a coefficient of performance (COP) for the entire waste heat process is usually assumed when converting to heating or cooling. Although this is not feasible for energy consisting of different qualities, in order to compare with the CEU model, the COPs for heating and cooling are assumed as 1.27 and 0.97, which are calculated:

$$COP_{heat} = (COP_{1AHP}\cdot 0.232 + \eta_{DH}\cdot 0.2)/0.432 \tag{28}$$

Table 2
The difference among the initial conditions of the four cases.

Item	Purchase heat	Sale heat	RC minTL	HST unit cost	Candidate technologies
Case A	Yes	No	400 °C	10$/kWh	All included
Case B	No	No	400 °C	10$/kWh	Without PV
Case C	Yes	No	510 °C	10$/kWh	With/without ORC
Case D	Yes	No	510 °C	5$/kWh	Without ORC

minTL means minimum temperature limit.

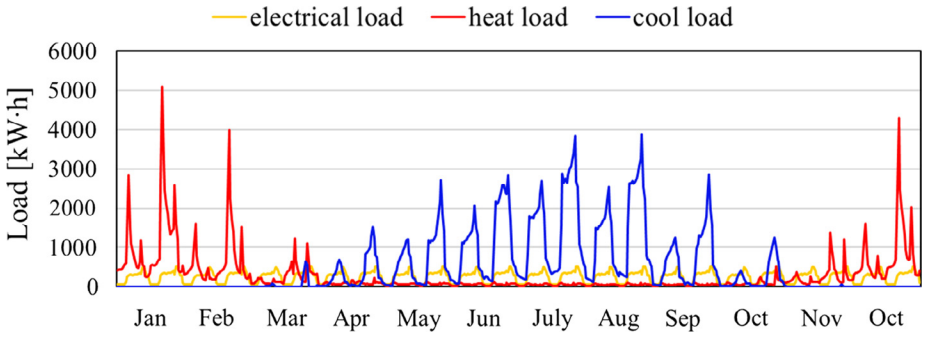

Fig. 3. The load profiles of the four cases.

$$COP_{cool} = (COP_{DARS}\cdot 0.232 + COP_{ARS}\cdot 0.2)/0.432 \quad (29)$$

The optimization results for the technologies selected and their capacities in the two model scenarios are shown in Table 5. For the CEU model in this case, heat cannot be sold, but can be purchased inexpensively, making electricity costs much more expensive than heating costs. Consequently, it is more economical to produce electricity with RC or ORC than by recovering exhaust heat; RC is selected owing to the sufficient exhaust temperature. Since the waste heat from jacket water cannot produce electricity, it is used to supply heating or cooling. It can be seen that the electricity produced by PV is minimal, because sunlight is not abundant and PV costs are relatively high. Since the exhaust temperature is high enough to drive DARS, it is configured to recover exhaust heat during cooling days; the electric chiller is used to supply cooling when there is not enough waste heat. The lowest exhaust temperature output from the RC is assumed to be 150 °C in Part 4, so that the remaining exhaust heat still can be recovered by the DH with cascade utilization. However, this segment of heat is small and the heat purchase price is low, so considering the cost of DH equipment, it is not economic to install. Figs. 4 and 5 describe the energy flow of the entire DES in the two typical days. Therein, the value in every energy flow is the sum of 24 h.

The WHRT selected significantly affects the efficiency of the entire system. However, the limits of traditional CCHP models preclude the optimization of WHRTs and efficient dispatch of waste heat utilization according to energy quality and load demand, which will be described in detail below. The objective function value of the traditional CCHP model is 205943.2, while that of the CEU model is 196236.1—almost 5% higher. Consequently, the CEU model can improve the profit of a DES and is more realistic for using waste heat of different quality in engineering practice.

Figs. 6 and 7 describe the electricity balance and heat balance of the CEU model during a typical day 1. During hours 1–5, the cost of electricity produced by the engine, including RC, is greater than the grid price. Additionally, the heat can be purchased at a low price, so it is not economical to produce heating with the waste heat. Therefore, the engine is shut down and the electricity and heating are supplied by the

Table 4
The main performance parameters of technologies.

Technology	Parameter value	
Photovoltaics	Solar efficiency: 0.15	Unit out ratio:0.15
Solar thermal	Solar efficiency: 0.69	Unit out ratio: 0.1
Internal combustion engine	Thermal efficiency: 0.4	
Natural gas boiler	Efficiency: 0.9	
Electric boiler	Efficiency: 0.9	
Natural gas chiller	COP: 1.2	
Electric chiller	COP: 5.6	
Heat pump	COP (heating): 3	COP (cooling): 5
Rankine cycle	Thermal efficiency: 0.18	
HT Organic Rankine Cycle	Thermal efficiency: 0.17	
LT Organic Rankine Cycle	Thermal efficiency: 0.08	
Double effect absorption refrigerator	COP: 1.2	
Absorption heat pump	COP: 1.6	
Single effect absorption refrigerator	COP: 0.7	
Direct heating	Efficiency: 0.9	
Electric battery	Charging efficiency: 0.9	Discharging efficiency: 0.9
	Max charge rate: 0.25	Max discharge rate: 0.25
	Min SOC: 0.1	
Heat storage tank	Charging efficiency: 0.9	Discharging efficiency: 0.8
	Max charge rate: 0.25	Max discharge rate: 0.25
	Min SOC: 0.15	
Cooling storage tank	Charging efficiency: 0.9	Discharging efficiency: 0.9
	Max charge rate: 0.25	Max discharge rate: 0.25
	Min SOC: 0	

Table 3
The main operative cost parameters.

Cost item	Value					
Nature gas price ($/kWh)	0.03					
Heat purchase price ($/kWh)	0.012					
Heat sale price ($/kWh)	0.010					
Time interval for electricity price	1th-5th	6th-8th	9th-11th	12th-18th	19th-21th	22th-24th
Electricity purchase price ($/kWh)	0.064	0.134	0.207	0.134	0.207	0.064
Electricity sale price ($/kWh)	0.054	0.114	0.176	0.114	0.176	0.054

Table 5
Optimization of equipment selection and capacity in Case A.

Technology	Capacity/kW (CEU)	Capacity/kW (CCHP)
ICE	960	Electricity 960, heat 1276, cooling 997
RC	87	/
DARS	668	/
JWDH	431	/
JWARS	335	/
EC	2868	2908
PV	392 (m^2)	392 (m^2)

grid and purchased heat, respectively. During hours 6–21, both the grid price and the electricity sale price are more expensive than the power generation cost of the engine, so the engine operates under the full load and all surplus electricity is sold. At the same time, all the exhaust heat is recovered by the RC to produce electricity. Although prices are low to purchases heat, jacket water waste heat is free, so it is used to supply heating and the remainder of needed heat is purchased. During hours 22–24, conditions are similar to hours 1–5; electricity is supplied from the grid and heating is purchased. The electricity balance of the traditional CCHP model is similar to Fig. 6, except for the lack of RC. However, it cannot optimize the best planning and operation schedule for WHRTs (Fig. 8), so all of the heat is used to supply heating. Any excess waste heat is discharged.

Figs. 9–11 describe the electricity, heating, and cooling balance of the CEU model during a typical day 2. In this load day, the most economical way to recover waste heat is to produce cooling. During hours 1–5 and 22–24, even though the waste heat is used up to produce cooling, the whole operation cost of the cogeneration system is larger than purchasing grid electricity to meet the electric load and drive the electrical chiller, owing to the low grid price. As a result, the engine is shut down and electricity, heating, and cooling are supplied by the grid, purchase, and electrical chiller, respectively. During hours 6–21, the engine runs under full load. According to optimization results, producing cooling is more economical than using grid electricity, so most of the time all of the exhaust waste heat and the jacket water waste heat are used to produce cooling. At hour 6, because the cooling load is small, part of the exhaust waste heat is used to produce electricity after satisfying the cooling load. The jacket water cannot drive RC to produce electricity, so it supplies heating. The conditions for hour 21 are similar. The most economical way to supply heating is to purchase as mentioned above, so at most of the time heat is purchased except for in hour 6 when jacket water heat is redundant to supply cooling. Fig. 12 is the distribution of the exhaust heat. During hours 6 and 21 when the cooling load is low, the exhaust drives the RC first and then the exhaust temperature is still high enough to drive the DARS. During hour 6, exhaust temperature coming out of the RC is 200 °C. Although the RC can continue to recover exhaust heat until the exhaust temperature reaches its lower limit (150 °C) and more jacket water is used to offer cooling, the remaining exhaust heat cannot drive the DARS: the exhaust heat below 150 °C will be wasted. Therefore, exhaust heat is only recovered above 200 °C, which represents the distinguishing energy level and segmented utilization of this optimization model. The electricity and heat balances of the traditional CCHP model are still similar with those of the CEU model, except for the lack of RC, while its cooling balance is simpler as shown in Fig. 13. When waste heat is in excess to produce cooling, the redundant heat is wasted instead of driving the RC.

5.2. Case B (waste heat sale)

In this case, conditions are the same as the base case, except that the heat cannot be purchased or sold, and the PV is removed. This case is aimed at studying the optimization results for different commercial modes. The optimization of equipment selection and capacity are shown in Table 6. It is expensive to supply heating using active technologies, such as natural gas boilers, owing to high natural gas prices, so 1AHP is selected and given priority to recover exhaust heat during the heating season. However, because the heat load is sometimes small and there is redundant waste heat, RC is selected to produce electricity

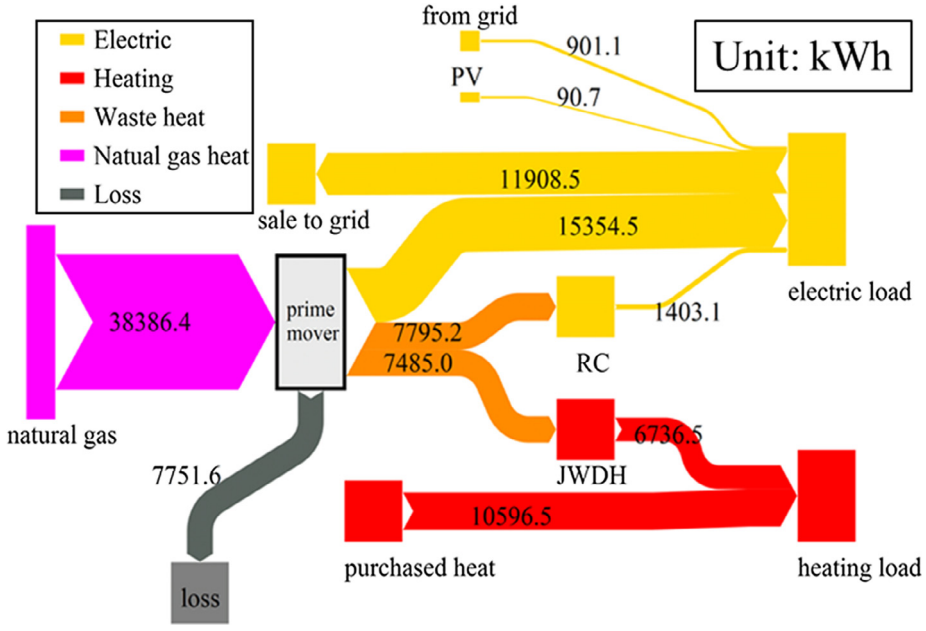

Fig. 4. Energy flow diagram for a typical winter (heating) day in Case A.

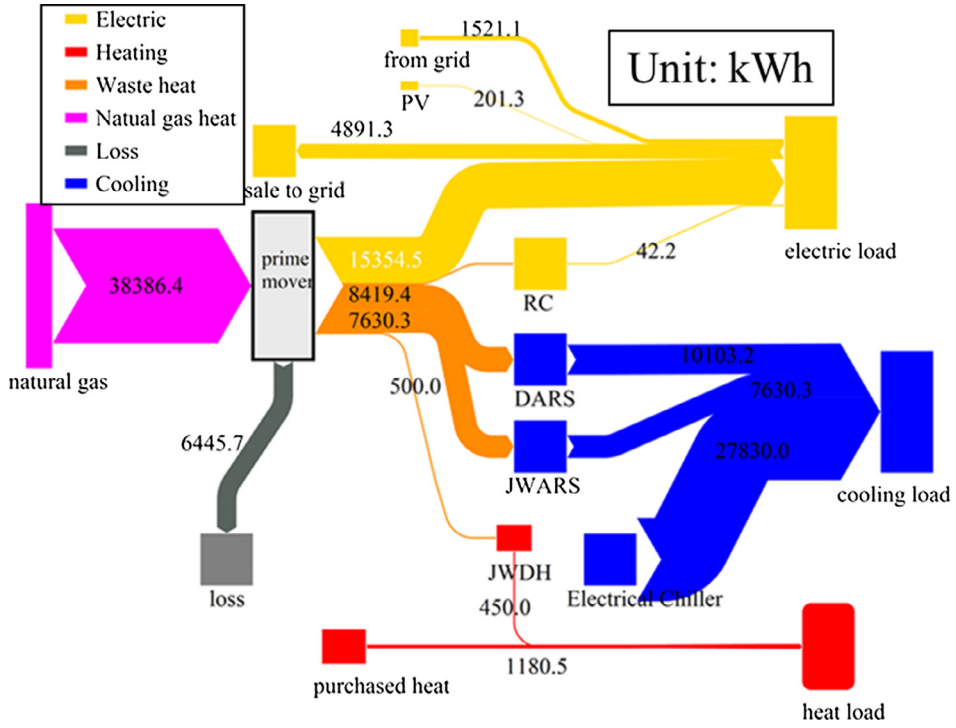

Fig. 5. Energy flow diagram for a summer (cooling) day in Case A.

with the waste heat in order to improve the whole energy utilization efficiency. During the cooling season, since the heat load is quite small, DARS technology is applied to recover exhaust waste heat. Jacket water waste heat only can be recovered by DH and ARS, so it is used to offer heating during the heating season and primarily supply cooling during the cooling season. According to optimization results, when waste heat is not sufficient, the natural gas boiler and electrical chiller are the best choices to supply heating and cooling.

Figs. 14 and 15 describe the electricity balance and heat balance in a typical day 1. During hours 1–5, despite the fact that grid electricity price (0.064$/kWh) is lower than the power generation cost of the engine (0.075//kWh), waste heat from exhaust and jacket water can meet the heat load, avoiding using the natural gas boiler with its expensive operation cost. Therefore, it is more economical to use the cogeneration system than offer heating and electricity separately. In order to meet the heat load by the waste heat, the extra electricity is sold to the grid. On the other hand, when the waste heat amount is enough to meet the heat load, the engine is shut down. Otherwise, the redundant waste heat only can be recovered by RC to produce

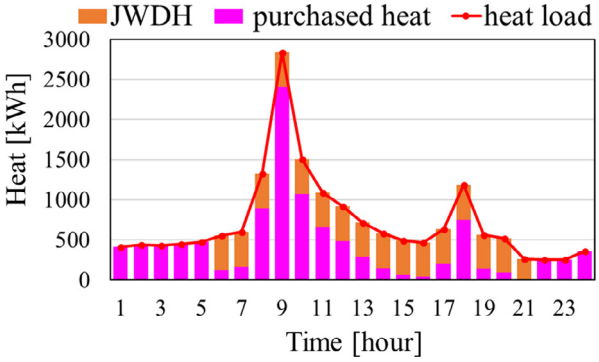

Fig. 7. Heat balance in the CEU model during a typical day 1 of Case A.

electricity; this is not the best strategy because of the low grid price and electricity sale price. During hours 6–21, since both the grid price and the electricity sale price are higher than the power generation cost of the engine, the engine operates under the full load and all the surplus

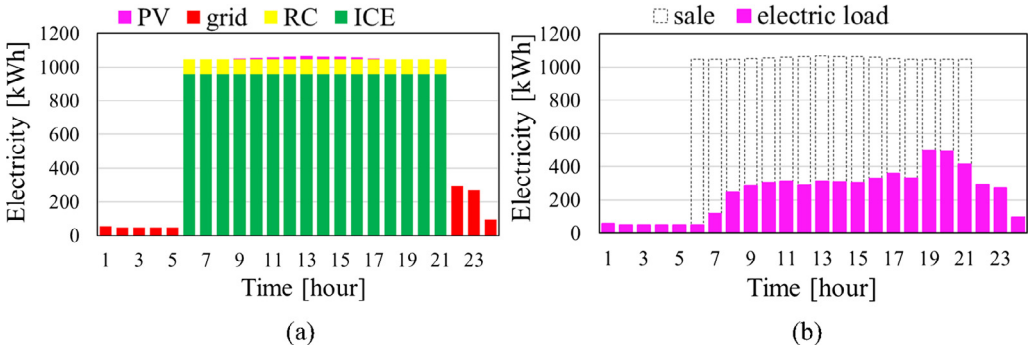

Fig. 6. Electricity balance in the CEU model during a typical day 1 of Case A (a. supply, b. demand).

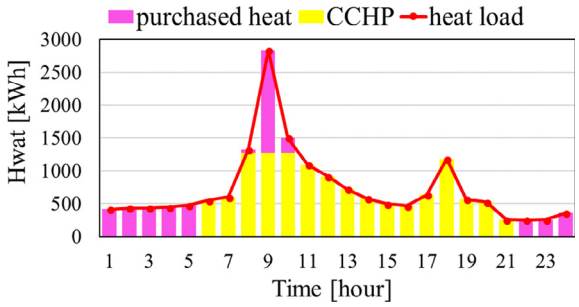

Fig. 8. Heat balance in the traditional CCHP model during a typical day 1 of Case A.

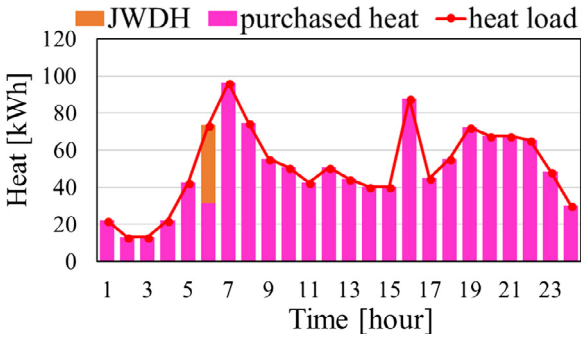

Fig. 10. Heat balance of the CEU model during a typical day 2 of Case A.

electricity is sold. As mentioned above, heat load takes top priority, but during this period the waste heat is redundant for supplying heating except during hours 9 and 10. Therefore, exhaust waste heat should be utilized in segments with a part of it recovered by the RC. Fig. 16 shows the distribution of exhaust energy utilization. It should be noted that, although the 1AHP has the priority, the RC needs high quality energy to drive, so exhaust heat should be first recovered by the RC (Fig. 16). In hours 9 and 10 waste heat is not sufficient, so all waste heat is used to supply heating using 1AHP and JWDH. The remaining heat load is supplied by the natural gas boiler. During hours 22–24, conditions are similar to hours 1–5.

Figs. 17–19 describe the electricity, heating, and cooling balance during a typical day 2. During a typical day 2, the heating load is met in priority and then the rest waste heat is used to supply cooling. During hours 1–5, the heat load is more economically provided by the cogeneration system as mentioned in typical day 1. Since the heat load is quite small, the engine runs under low load as well. Therefore, electricity generated by the engine is not sufficient and the grid supplies the remainder of needed electricity. Because the amount of waste heat is small and because supplying heating is the priority, cooling is primarily generated using the electric chiller. During hours 6–21, the engine operates under the full load owing to the low grid price. After satisfying the small heat load, most of the waste heat is used to produce cooling. In hour 6, because both the heat and cooling loads are small, extra waste heat is recovered by the RC. In the other hours, there is not enough waste heat to meet the cooling load, so the electrical chiller is used. Conditions of hours 22–24 are similar to hours 1–5. It should be noted that, at most times, the heating load is supplied by jacket water heat with low energy quality so that more exhaust heat can drive the DARS with higher energy quality requirements and better efficiency. This shows the model's ability to distinguishing energy levels and cascade utilization as well.

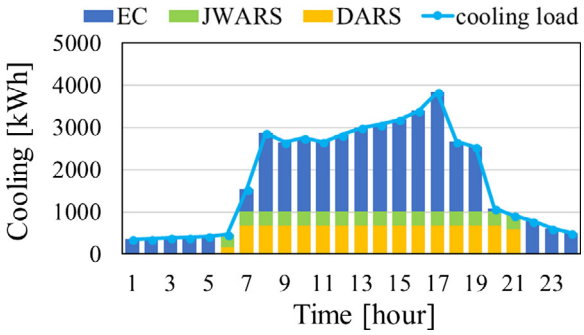

Fig. 11. Cooling balance of the CEU model during a typical day 2 of Case A.

5.3. *Case C (high lower limit temperature of RC)*

Conditions are the same with base Case A, except that the lowest temperature to drive the RC is assumed to be 510 °C. Case C is aimed to prove that this optimization model can screen out the technology that cannot be driven by the heat source. Since the exhaust temperature is not high enough to drive RC, there is no RC in the optimization results as shown in Table 7. By contrast, ORC is a suitable technology to recover low temperature waste heat as mentioned in Section 3, so it is selected in the optimization results. Since both the ORC and RC are used to offer electricity, and other conditions are the same as in the base case, operation is also nearly the same as the base case. It is not shown again here.

If the ORC is not the candidate technology (because of some reason such as user preference), exhaust waste heat only can be recovered by 1AHP or DH during the heating season. Then the optimization results for this case are shown in Table 8. Although the 1AHP can produce much more heat than the DH, its investment and operating costs are

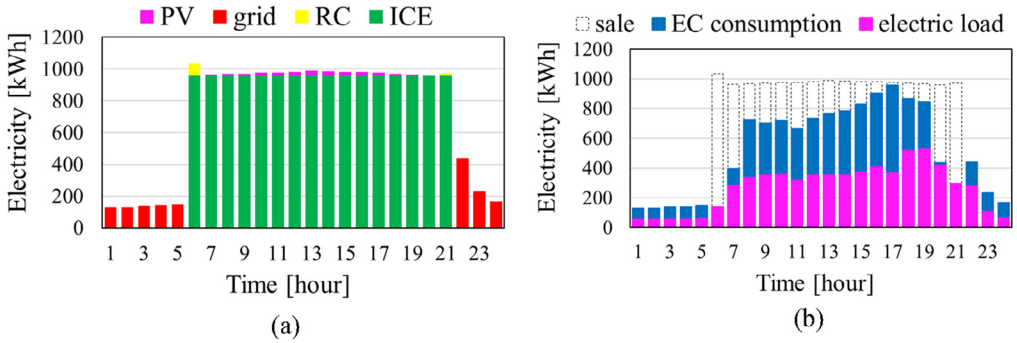

Fig. 9. Electricity balance of the CEU model during a typical day 2 of Case A (a. supply, b. demand).

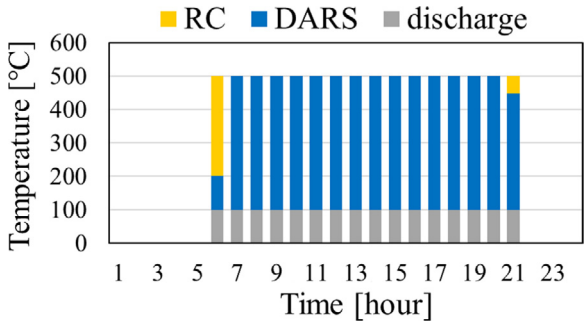

Fig. 12. Exhaust energy distribution during typical day 2 of Case A.

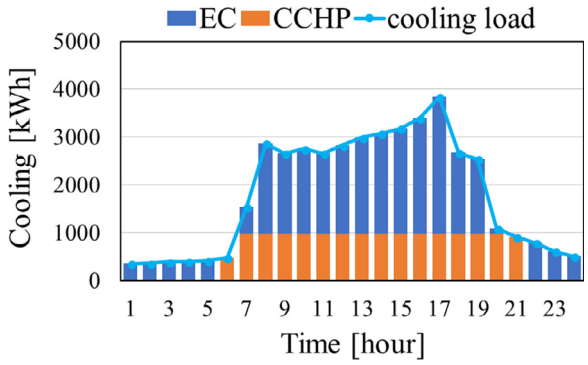

Fig. 13. Cooling balance of the CCHP model during a typical day 2 of Case A.

Table 6
Optimization of equipment selection and capacity in Case B.

Technology	Capacity (kW)
ICE	960
RC	87
1AHP	890
DARS	668
JWDH	431
JWARS	335
NB	3370
EC	2868

greater. Additionally, heat purchase price is low, leading to a low profit of supply heating by waste heat. As a result, the less expensive DH is the optimal choice. The other equipment and their capacities are the same as the base case.

Except for the heat balance during typical day 1, other energy balances without ORC are quite similar to those in the base case, so they are not described in detail here. Fig. 20 describes the heat balance during a typical day 1. During hours 1–5 and 22–24, the engine is shut down for the same reason as the base case. The electricity and heating are supplied by the grid and purchasing heat, respectively. During hours 6–21, the engine runs under full load. Both exhaust and jacket water waste heat are used to supply heating. When the heat load is small, there is redundant waste heat, but it cannot be sold or stored, so it is discharged. As shown in Fig. 21, the final discharge exhaust temperature is periodically much higher than 100 °C.

5.4. Case D (with heat storage)

The former Case C shows the necessity of heat storage. In order to use up the waste heat, the unit investment cost of the heat tank is assumed to be ($5/kWh). Consequently, the heat tank appears in the optimization results; its capacity is 3647 kW. Other equipment is the same as that in Case C. Because the other initial conditions of cases D and C are the same, the operation of cooling and electricity balance is the same as well, and they are not shown again. In this case, only heat balance is analyzed.

As shown in Fig. 22, during a typical day 1, during the hours 6–7, 13–17, and 19–21 when the waste heat is greater than the heat load, the redundant segment is stored in the heat tank. Therefore, exhaust heat and jacket water heat can be fully used. The stored heat releases to supply heating when the waste heat is not sufficient, which can reduce the amount of purchased heat. Since there is a limit on the maximum discharge rate of the heat tank, the stored heat sometimes cannot meet the heat demand alone; purchased heat is required (such as at hours 8–10). Similarly, during a typical day 2 as shown in Fig. 23, in hours 6 and 21 when the cooling load is small, the redundant waste heat is stored and releases at another time.

5.5. Summary of cases

The model presented in this paper is intended to help actual heat utilization engineering project design and operation. Real building load data are used for validation the model's capability. Four case scenarios are studied. The base Case A indicates that the DES for a real demand load can use the model to configure the optimal waste heat recovery technologies from a variety of candidates, as well as other equipment. Because of the energy equality distinguishing function, this model can help users to plan and operate a more efficient and cost-effective DES than the traditional CCHP model. Case B shows that the model gives users a totally different result when the business model changes, especially a very detailed dispatch of waste heat segmented utilization,

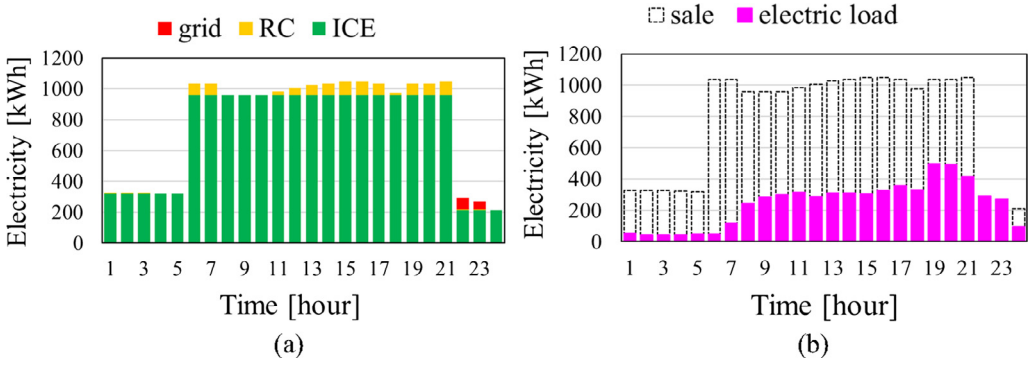

Fig. 14. Electricity balance during a typical day 1 in Case B (a. supply, b. demand).

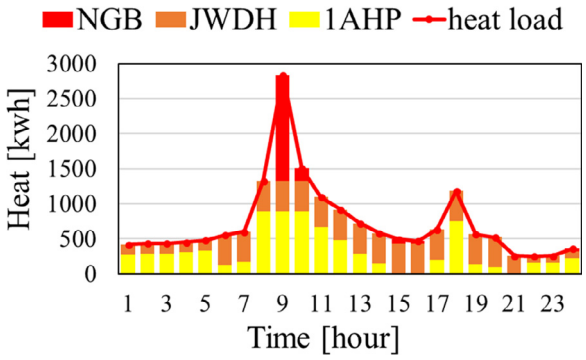

Fig. 15. Heat balance during a typical day 1 in Case B.

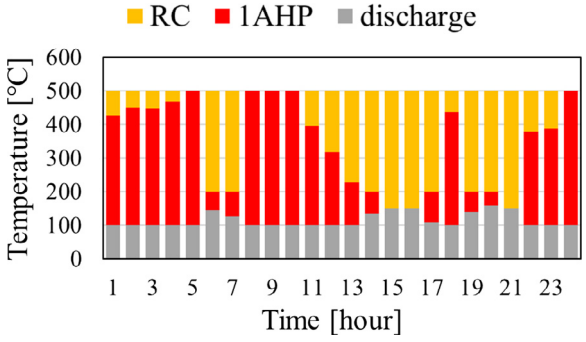

Fig. 16. Exhaust energy distribution during a typical day 1 in Case B.

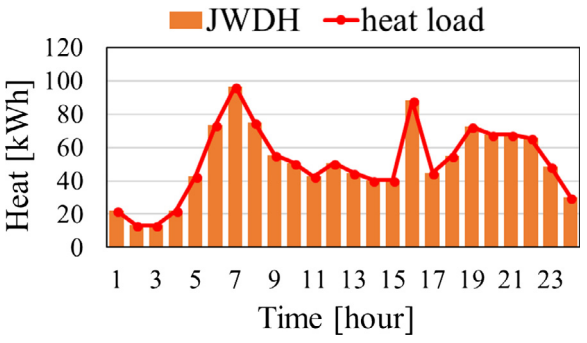

Fig. 18. Heat balance during a typical day 2 in Case B.

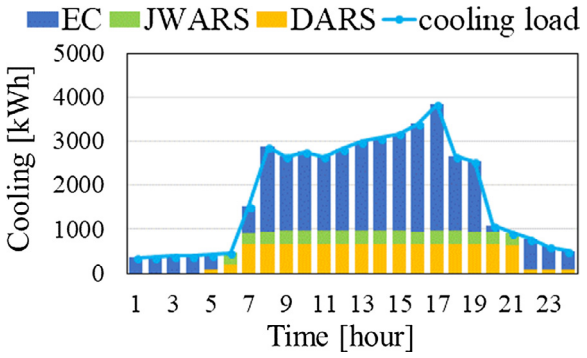

Fig. 19. Cooling balance during a typical day 2 in Case B.

which is a good reference to operate the real DES and make it more efficient. Case C proves the function of screening out the technologies that cannot be driven by the heat source to make the optimization results more reasonable in engineering practices. Finally, Case D shows that the energy storage modules can be included in the results when necessary. Based on the case study, even though the model is developed for multi-segment waste heat utilization, results show that a DES may only use two or three temperature segments. The conclusion on system configuration is in line with actual engineering project experience.

Table 7
Optimization of equipment selection and capacity in Case C.

Technology	Capacity (kW)
ICE	960
HT-ORC	82
DARS	668
JWDH	431
JWARS	335
PV	392 (m^2)
EC	2868

6. Conclusion

Waste heat is a common energy type in many industrial application and distributed energy systems. Utilizing waste heat efficiently is an important research topic with great potential for energy and cost savings. Several energy technologies can function under different waste heat inputs to produce electricity, heating, or cooling. In order to distinguish different waste heat thermal energy quality in distributed

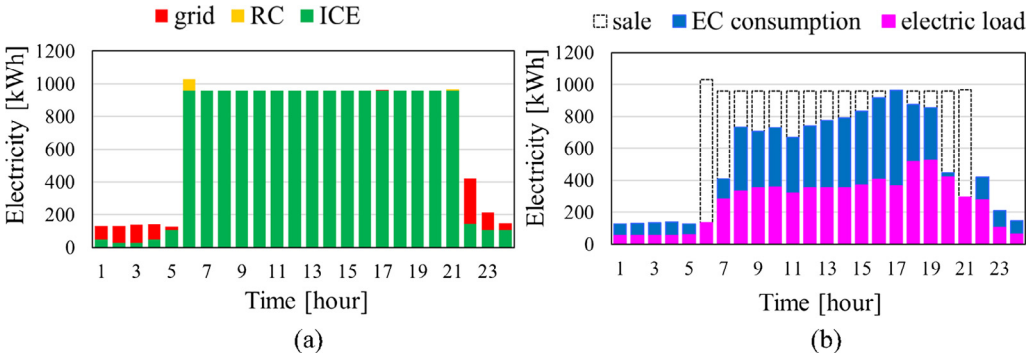

Fig. 17. Electricity balance during a typical day 2 in Case B (a. supply, b. demand).

Table 8
Optimization of equipment selection and capacity in Case C without ORC.

Technology	Capacity (kW)
ICE	960
DH	501
DARS	668
JWDH	431
JWARS	335
PV	392 (m^2)
EC	2868

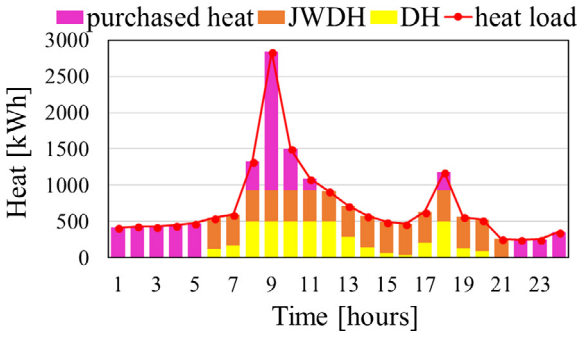

Fig. 20. Heat balance during a typical day 1 in Case C.

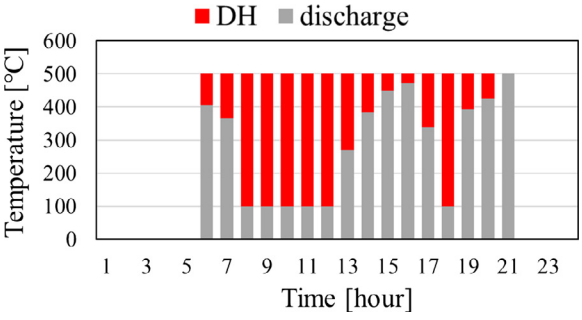

Fig. 21. Exhaust energy distribution during a typical day 1 in Case C without ORC.

energy system planning and operation optimization, a novel and detailed mixed integer linear programming optimization model based on energy cascade utilization is proposed in this study. This study defines an energy flow tower and setup operation temperature thresholds for each technology. The developed model is validated in four typical cases and compared with the traditional model where different waste heat temperature grades are not considered.

The developed model can distinguish energy quality and perform a detailed optimal selection of waste heat recovery technologies for different types of heat sources based on the load demand and energy quality. This function can avoid unreasonable energy upgrade utilization and efficiency loss due to energy degrade utilization, making the optimization results more realistic and efficient. Furthermore, the model can give a detailed operation schedule of different waste heat recovery technologies by segmented utilization, making full use of waste heat. Because of these advantages, the model can increase energy cost savings by around 5% compared with traditional models with no temperature grade differentiation. This amount of energy saving should be interpreted as a realistic estimation of the additional energy saving due to energy cascade utilization based on a technologically constrained model. Compared to previous works, the proposed method can fully leverage the energy cascade mechanism by putting forward a detailed investment and dispatch strategy.

In summary, the model developed in this study can produce more reasonable, economical and efficient optimization results for the configuration and operation of a distributed energy system by using waste heat based on energy equality. In the future, ongoing research will consider the part-load performance of equipment to make this model more practical for engineering applications. Additionally, high temperature heat storage will be added to the model to improve waste heat utilization during the operation period as well as other types of heat source with the proposed method of energy tower.

Acknowledgements

Lawrence Berkeley National Laboratory was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and Energy Foundation China. Support also came from Chinese researchers in many research institutes who provided materials, guidance, and advice.

This work was supported by the State Key Program of National Natural Science Foundation of China (No. 51636005).

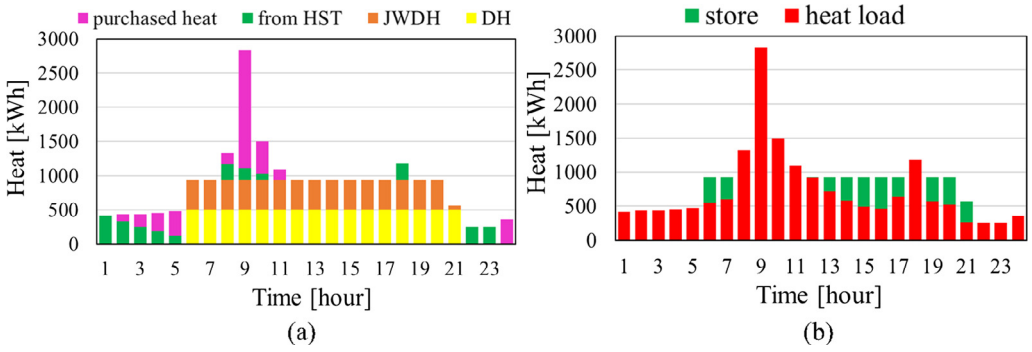

Fig. 22. Heat balance during a typical day 1 in Case D (a. supply, b. demand).

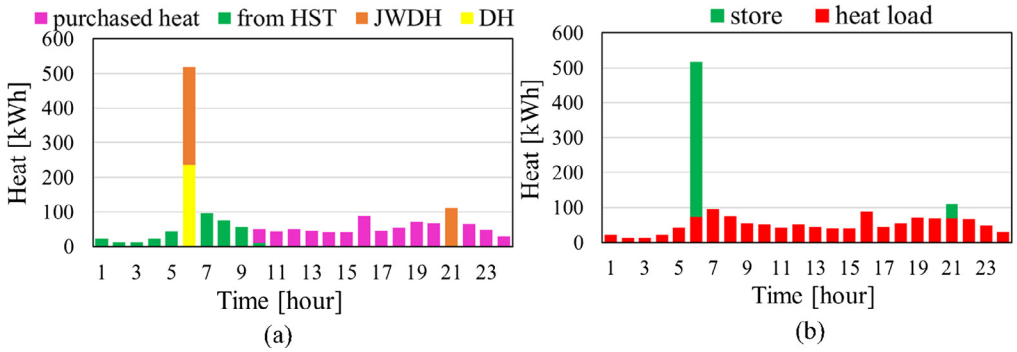

Fig. 23. Heat balance during a typical day 2 in Case C (a. supply, b. demand).

Appendix A

A.1. Exhaust heat recovery by double-effect absorption refrigeration system (DARS)

Energy balance of DARS:

$$Q_{m,d,h}^{DARS} = Z_{m,d,h}^{DARS}(T_{m,d,h}^{ex3} - T_{m,d,h}^{ex3'})m_{m,d,h}^{ex}C_p^{ex} \tag{A1}$$

$$QC_{m,d,h}^{DARS} = Q_{m,d,h}^{DARS}COP^{DARS} \tag{A2}$$

Capacity constraints of DARS:

$$Cp_{\min}^{DARS}X^{DARS} \leqslant Cp_{rat}^{DARS} \leqslant Cp_{\max}^{DARS}X^{DARS} \tag{A3}$$

Temperature constraints of DARS:

$$\begin{cases} Z_{m,d,h}^{DARS} \in [0, 1] \\ T_{m,d,h}^{ex3} \geqslant Z_{m,d,h}^{DARS}200 \\ T_{m,d,h}^{ex3} \leqslant Z_{m,d,h}^{DARS}(T_{m,d,h}^{ex2'} - 200) + 200 \end{cases} \tag{A4}$$

$$T_{m,d,h}^{ex3} \geqslant T_{m,d,h}^{ex3'} \geqslant Z_{m,d,h}^{DARS}100 \tag{A5}$$

A.2. Exhaust heat recovery by absorption heat pump of the first type (1AHP)

Energy balance of 1AHP:

$$Q_{m,d,h}^{1AHP} = Z_{m,d,h}^{1AHP}(T_{m,d,h}^{ex4} - T_{m,d,h}^{ex4'})m_{m,d,h}^{ex}C_p^{ex} \tag{A6}$$

$$H_{m,d,h}^{1AHP} = Q_{m,d,h}^{1AHP}COP^{1AHP} \tag{A7}$$

Capacity constraints of 1AHP:

$$Cp_{\min}^{1AHP}X^{1AHP} \leqslant Cp_{rat}^{1AHP} \leqslant Cp_{\max}^{1AHP}X^{1AHP} \tag{A8}$$

Temperature constraints of 1AHP:

$$\begin{cases} Z_{m,d,h}^{1AHP} \in [0, 1] \\ T_{m,d,h}^{ex4} \geqslant Z_{m,d,h}^{1AHP}200 \\ T_{m,d,h}^{ex4} \leqslant Z_{m,d,h}^{1AHP}(T_{m,d,h}^{ex3'} - 200) + 200 \end{cases} \tag{A9}$$

$$T_{m,d,h}^{ex4} \geqslant T_{m,d,h}^{ex4'} \geqslant Z_{m,d,h}^{1AHP}100 \tag{A10}$$

A.3. Exhaust heat recovery by low temperature y(LT ORC)

Energy balance of LT ORC:

$$Q_{m,d,h}^{LTORC} = Z_{m,d,h}^{LTORC}(T_{m,d,h}^{ex5} - T_{m,d,h}^{ex5'})m_{m,d,h}^{ex}C_p^{ex} \tag{A11}$$

$$P_{m,d,h}^{LTORC} = Q_{m,d,h}^{LTORC}\eta^{LTORC} \tag{A12}$$

Capacity constraints of LT ORC:

$$Cp_{\min}^{LTORC}X^{LTORC} \leqslant Cp_{rat}^{LTORC} \leqslant Cp_{\max}^{LTORC}X^{LTORC} \tag{A13}$$

Temperature constraints of LT ORC:

$$\begin{cases} Z_{m,d,h}^{LTORC} \in [0,\ 1] \\ T_{m,d,h}^{ex5} \geqslant Z_{m,d,h}^{LTORC} 130 \\ T_{m,d,h}^{ex5} \leqslant Z_{m,d,h}^{LTORC} (T_{m,d,h}^{ex4'} - 130) + 130 \end{cases} \tag{A14}$$

$$T_{m,d,h}^{ex5} \geqslant T_{m,d,h}^{ex5'} \geqslant Z_{m,d,h}^{LTORC} 100 \tag{A15}$$

A.4. Exhaust heat recovery of absorption refrigeration system (ARS)

Energy balance of ARS:

$$\mathrm{Q}_{m,d,h}^{ARS} = Z_{m,d,h}^{ARS} (T_{m,d,h}^{ex6} - T_{m,d,h}^{ex6'}) m_{m,d,h}^{ex} C_p^{ex} \tag{A16}$$

$$QC_{m,d,h}^{ARS} = \mathrm{Q}_{m,d,h}^{ARS} COP^{ARS} \tag{A17}$$

Capacity constraints of ARS:

$$Cp_{\min}^{ARS} X^{ARS} \leqslant Cp_{rat}^{ARS} \leqslant Cp_{\max}^{ARS} X^{ARS} \tag{A18}$$

Temperature constraints of ARS:

$$\begin{cases} Z_{m,d,h}^{ARS} \in [0,\ 1] \\ T_{m,d,h}^{ex6} \geqslant Z_{m,d,h}^{ARS} 110 \\ T_{m,d,h}^{ex6} \leqslant Z_{m,d,h}^{ARS} (T_{m,d,h}^{ex5'} - 110) + 110 \end{cases} \tag{A19}$$

$$T_{m,d,h}^{ex6} \geqslant T_{m,d,h}^{ex6'} \geqslant Z_{m,d,h}^{ARS} 100 \tag{A20}$$

A.5. Exhaust waste heat recovery by direct heating (DH)

Energy balance of DH:

$$Q_{m,d,h}^{DH} = Z_{m,d,h}^{DH} (T_{m,d,h}^{ex7} - T_{m,d,h}^{ex7'}) m_{m,d,h}^{ex} C_p^{ex} \tag{A21}$$

$$H_{m,d,h}^{DH} = \mathrm{Q}_{m,d,h}^{DH} \eta^{DH} \tag{A22}$$

Capacity constraints of DH:

$$Cp_{\min}^{DH} X^{DH} \leqslant Cp_{rat}^{DH} \leqslant Cp_{\max}^{DH} X^{DH} \tag{A23}$$

$$\begin{cases} Z_{m,d,h}^{DH} \in [0,\ 1] \\ T_{m,d,h}^{ex7} \geqslant Z_{m,d,h}^{DH} 70 \\ T_{m,d,h}^{ex7} \leqslant Z_{m,d,h}^{DH} (T_{m,d,h}^{ex6'} - 70) + 70 \end{cases} \tag{A24}$$

Temperature constraints of DH:

$$T_{m,d,h}^{ex7} \geqslant T_{m,d,h}^{ex7'} \geqslant Z_{m,d,h}^{DH} 100 \tag{A25}$$

References

[1] Ren Hongbo, Gao Weijun. A MILP model for integrated plan and evaluation of distributed energy systems. Appl Energy 2010;87(3):1001–14.
[2] Akorede Mudathir Funsho, Hizam Hashim, Pouresmaeil Edris. Distributed energy resources and benefits to the environment. Renew Sustain Energy Rev 2010;14.2:724–34.
[3] Gu Wei, et al. Modeling planning and optimal energy management of combined cooling, heating and power microgrid: a review. Int J Electr Power Energy Syst 2014;54:26–37.
[4] Al Moussawi Houssein, Fardoun Farouk, Louahlia Hasna. Selection based on differences between cogeneration and trigeneration in various prime mover technologies. Renew Sustain Energy Rev 2017;74:491–511.
[5] Forman Clemens, et al. Estimating the global waste heat potential. Renew Sustain Energy Rev 2016;57:1568–79.
[6] Wang Yaodong, et al. An investigation of a household size trigeneration running with hydrogen. Appl Energy 2011;88(6):2176–82.
[7] Cao Tao, Hwang Yunho, Radermacher Reinhard. Development of an optimization based design framework for microgrid energy systems. Energy 2017;140:340–51.
[8] Barelli L, et al. Dynamic analysis of PEMFC-based CHP systems for domestic application. Appl Energy 2012;91(1):13–28.
[9] Lecompte Steven, et al. Part load based thermo-economic optimization of the Organic Rankine Cycle (ORC) applied to a combined heat and power (CHP) system. Appl Energy 2013;111:871–81.
[10] Destro Nicola, et al. Components design and daily operation optimization of a hybrid system with energy storages. Energy 2016;117:569–77.
[11] Alvarado Dagoberto Cedillos, et al. A Technology Selection and Operation (TSO) optimisation model for distributed energy systems: Mathematical formulation and case study. Appl Energy 2016;180:491–503.
[12] Stoppato Anna, et al. A model for the optimal design and management of a co-generation system with energy storage. Energy Build 2016;124:241–7.
[13] Al Moussawi Houssein, Fardoun Farouk, Louahlia-Gualous Hasna. Review of tri-generation technologies: design evaluation, optimization, decision-making, and selection approach. Energy Convers Manage 2016;120:157–96.
[14] Liu Mingxi, Shi Yang, Fang Fang. Optimal power flow and PGU capacity of CCHP systems using a matrix modeling approach. Appl Energy 2013;102:794–802.
[15] Mago Pedro J, Hueffed Anna K. Evaluation of a turbine driven CCHP system for large office buildings under different operating strategies. Energy Build 2010;42(10):1628–36.
[16] Oh Si-Doek, et al. Optimal operation of a 1-kW PEMFC-based CHP system for residential applications. Appl Energy 2012;95:93–101.
[17] Basrawi Firdaus, Yamada Takanobu, Obara Shin'ya. Economic and environmental based operation strategies of a hybrid photovoltaic–microgas turbine trigeneration system. Appl Energy 2014;121:174–83.
[18] Marler R Timothy, Arora Jasbir S. Survey of multi-objective optimization methods

for engineering. Struct Multidiscipl Optim 2004;26(6):369–95.
[19] Hajabdollahi Hassan. Investigating the effects of load demands on selection of optimum CCHP-ORC plant. Appl Therm Eng 2015;87:547–58.
[20] Velik Rosemarie, Nicolay Pascal. Grid-price-dependent energy management in microgrids using a modified simulated annealing triple-optimizer. Appl Energy 2014;130:384–95.
[21] Sayyaadi Hoseyn, Babaie Meisam, Farmani Mohammad Reza. Implementing of the multi-objective particle swarm optimizer and fuzzy decision-maker in exergetic, exergoeconomic and environmental optimization of a benchmark cogeneration system. Energy 2011;36(8):4777–89.
[22] Chen XP, et al. Dynamic programming for optimal operation of a biofuel micro CHP-HES system. Appl Energy 2017;208:132–41.
[23] Celli G, et al. Optimal participation of a microgrid to the energy market with an intelligent EMS. Power Engineering Conference, 2005. IPEC 2005. The 7th International. IEEE; 2005.
[24] Niknam Taher, Azizipanah-Abarghooee Rasoul, Narimani Mohammad Rasoul. An efficient scenario-based stochastic programming framework for multi-objective optimal micro-grid operation. Appl Energy 2012;99455.
[25] Ommen Torben, Markussen Wiebke Brix, Elmegaard Brian. Comparison of linear, mixed integer and non-linear programming methods in energy system dispatch modelling. Energy 2014;74:109–18.
[26] Jin Ming, et al. Microgrid to enable optimal distributed energy retail and end-user demand response. Appl Energy 2018;210:1321–35.
[27] He, Xian, et al. Shift, not drift: towards active demand response and beyond; 2013.
[28] Brusco Giovanni, et al. Energy management system for an energy district with demand response availability. IEEE Trans Smart Grid 2014;5(5):2385–93.
[29] Dufo-López Rodolfo. Optimisation of size and control of grid-connected storage under real time electricity pricing conditions. Appl Energy 2015;140:395–408.
[30] Jin Ming, et al. MOD-DR microgrid optimal dispatch with demand response. Appl Energy 2017;187:758–76.
[31] Aghaali Habib, Ångström Hans-Erik. A review of turbocompounding as a waste heat recovery system for internal combustion engines. Renew Sustain Energy Rev 2015;49:813–24.
[32] Wang Tianyou, et al. A review of researches on thermal exhaust heat recovery with Rankine cycle. Renew Sustain Energy Rev 2011;15(6):2862–71.
[33] Ru-mou LIN, Hong-guang JIN, Rui-xian CAI. Integrated energy system of gas turbine and cascade utilization of thermal energy. Gas Turbine Technol 2008;1:002.
[34] Yang Mina, et al. High efficiency H2O/LiBr double effect absorption cycles with multi-heat sources for tri-generation application. Appl Energy 2017;187:243–54.
[35] Mohammadi Amin, et al. Thermodynamic analysis of a combined gas turbine, ORC cycle and absorption refrigeration for a CCHP system. Appl Therm Eng 2017;111:397–406.
[36] Han Wei, et al. New hybrid absorption–compression refrigeration system based on cascade use of mid-temperature waste heat. Appl Energy 2013;106:383–90.
[37] Jing Xuye, Zheng Danxing. Effect of cycle coupling-configuration on energy cascade utilization for a new power and cooling cogeneration cycle. Energy Convers Manage 2014;78:58–64.
[38] Chen Min, et al. Energy efficiency analysis and impact evaluation of the application of thermoelectric power cycle to today's CHP systems. Appl Energy 2010;87(4):1231–8.
[39] Bischi Aldo, et al. A detailed MILP optimization model for combined cooling, heat and power system operation planning. Energy 2014;74:12–26.
[40] Gao Penghui, et al. Energy matching and optimization analysis of waste to energy CCHP (combined cooling, heating and power) system with exergy and energy level. Energy 2015;79:522–35.
[41] Liang Youcai, et al. Analysis of an electricity–cooling cogeneration system based on RC–ARS combined cycle aboard ship. Energy Convers Manage 2013;76:1053–60.
[42] Kehlhofer Rolf, et al. Combined-cycle gas & steam turbine power plants. Pennwell Books; 2009.
[43] Gewald Daniela, et al. Waste heat recovery from a landfill gas-fired power plant. Renew Sustain Energy Rev 2012;16(4):1779–89.
[44] Pezzuolo Alex, et al. The ORC-PD: a versatile tool for fluid selection and Organic Rankine Cycle unit design. Energy 2016;102:605–20.
[45] Benato Alberto. Improving the efficiency of a cataphoresis oven with a cogenerative organic Rankine cycle unit. Therm Sci Eng Prog 2018;5:182–94.
[46] Shu Gequn, et al. Alkanes as working fluids for high-temperature exhaust heat recovery of diesel engine using organic Rankine cycle. Appl Energy 2014;119:204–17.
[47] Gadd Henrik, Werner Sven. Achieving low return temperatures from district heating substations. Appl Energy 2014;136:59–67.
[48] Li meng. Comparative study on heat pump technology based on waste heat recovery. Tianjin University, master thesis; 2013 [in Chinese].
[49] Wang xin. Research and application of waste heat utilization of thermal power unit based on absorption heat pump. North China electric power university, master thesis; 2015.
[50] Zhang Xiaodong, Dapeng Hu. Performance analysis of the single-stage absorption heat transformer using a new working pair composed of ionic liquid and water. Appl Therm Eng 2012;37:129–35.
[51] Wu DaWei, Wang RuZhu. Combined cooling, heating and power: a review. Prog Energy Combust Sci 2006;32(5–6):459–95.
[52] Deng J, Wang RZ, Han GY. A review of thermally activated cooling technologies for combined cooling, heating and power systems. Prog Energy Combust Sci 2011;37(2):172–203.
[53] Ommen Torben, Markussen Wiebke Brix, Elmegaard Brian. Lowering district heating temperatures – impact to system performance in current and future Danish energy scenarios. Energy 2016;94:273–91.
[54] Li You-Rong, Wang Jian-Ning, Mei-Tang Du. Influence of coupled pinch point temperature difference and evaporation temperature on performance of organic Rankine cycle. Energy 2012;42(1):503–9.
[55] Wang Xuan, et al. Engine working condition effects on the dynamic response of organic Rankine cycle as exhaust waste heat recovery system. Appl Therm Eng 2017;123:670–81.
[56] Shu Gequn, Wang Xuan, Tian Hua. Theoretical analysis and comparison of Rankine cycle and different organic Rankine cycles as waste heat recovery system for a large gaseous fuel internal combustion engine. Appl Therm Eng 2016;108:525–37.
[57] Kialashaki Yaser. A linear programming optimization model for optimal operation strategy design and sizing of the CCHP systems. Energy Efficiency 2018;11(1):225–38.